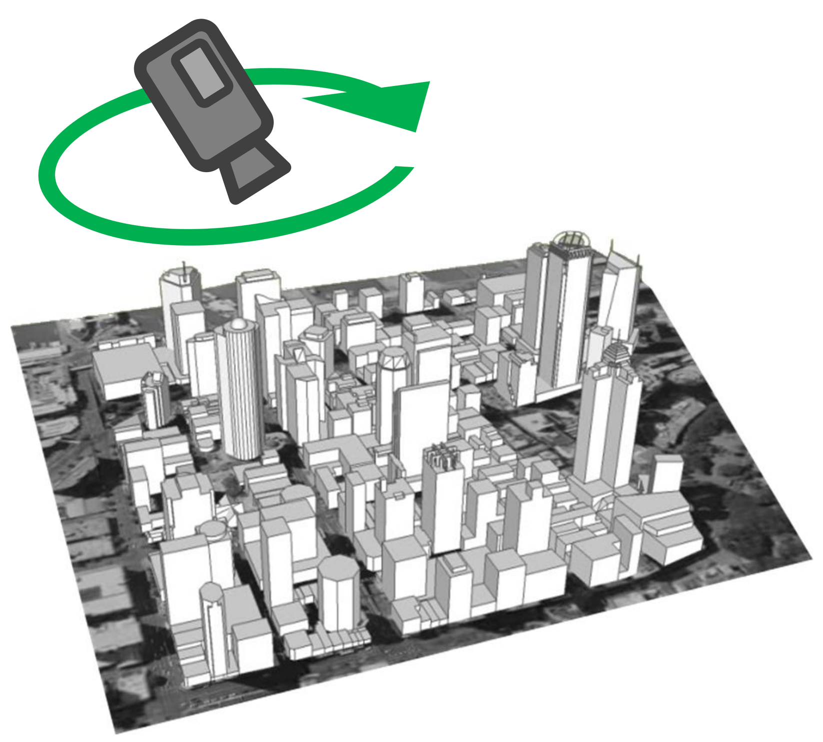
Pixel-wise Motion Detection in Persistent Aerial Video Surveillance

Grace Vesom <grace@llnl.gov>

Introduction

Aerial Video Surveillance (AVS) is employed over wide area, for hours at a time resulting in immense data collections.



Post ground stabilization, structural objects with depth appear to have precessive motion due to sensor movement alongside objects undergoing true, independent motion in the scene.

Computational objective

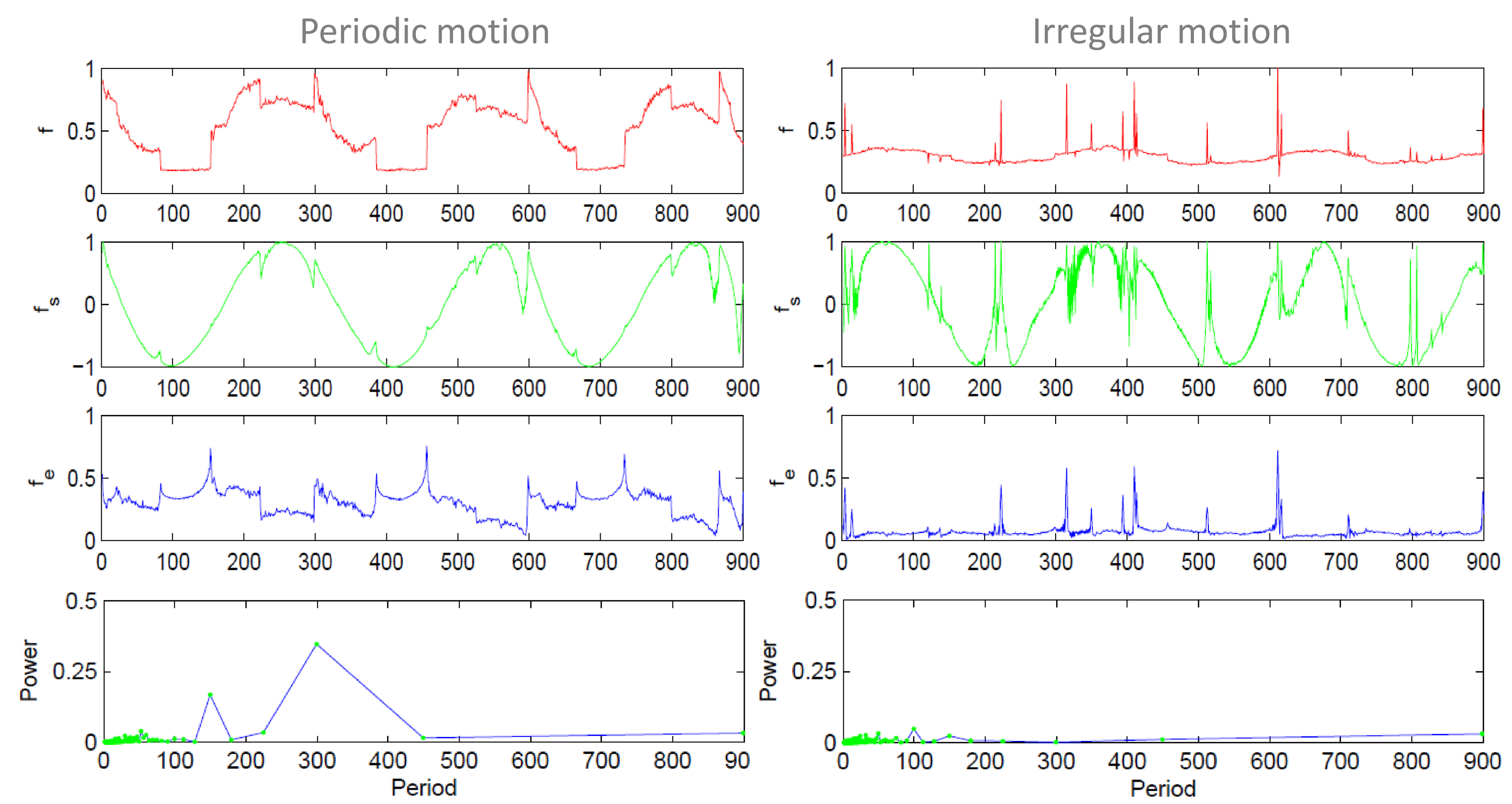
Disambiguate independent and structural motion in Wide Area Motion Imagery (WAMI) efficiently and robustly.

Applications

- Increase video compression rate via object-oriented video compression
- Robustly identify moving objects (vehicles, persons, sudden change in scenery) for end users in what can often be sparse terrain

Computational view of data

On a pixel level, the structural motion appears periodic, while the true, independent motion appears to have an irregular pattern over time.



The normalized signal (red) can be decomposed using Owen's equation $f(\mathbf{x}) = f_s(\mathbf{x}) \odot f_e(\mathbf{x})$ into a smooth (green) and local energy (blue) component. A metric is derived via time-frequency analysis (last row power spectrum) on the local energy component which further allows parameterized motion classification.

Why does Owen's equation apply here? *Phase congruency*.

Phase congruency (PC) is maximal when a signal's Fourier decomposition components align in phase and is employed as a low-level salient feature. Studies on PC are motivated by early human vision, but intended for application in computer vision. Our interest lies in its use in the single temporal direction, where the utility of visual saliency still applies, and we seek changes in the raw data that would be comparable to a visual shift. *The PC of a 1D signal is directly proportional to its local energy* $\mathrm{PC}(\mathbf{x}) \propto f_e(\mathbf{x})$.

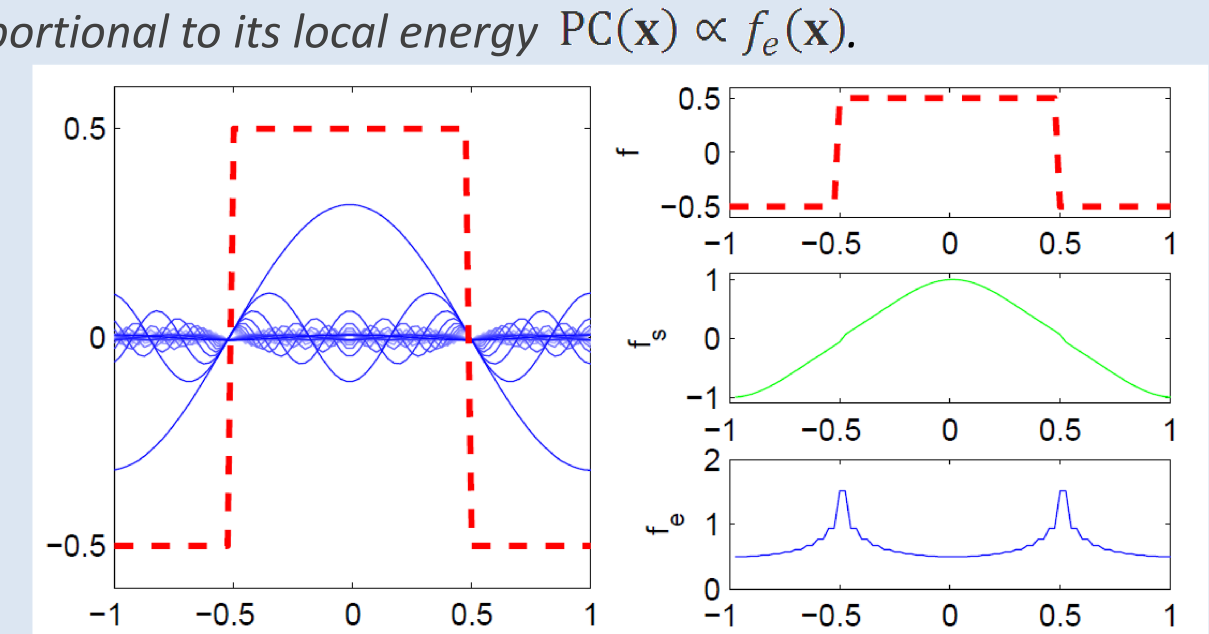


Figure 1. Square wave and its signal decomposition. (a) The Fourier decomposition of the square wave. PC is maximal at step edges. (b) The decomposition of the square wave via the $\odot$ operator, notice the locations of zero-crossings of the square wave and the peaks in $f_e(\mathbf{x})$.

Signal decomposition

Let $f(\mathbf{x})$ be the intensity normalized 1D vector for a fixed time period n at location $\mathbf{x} = (x, y)$. Owen's equation for signal decomposition is as follows:

$$\begin{aligned} f(\mathbf{x}) &= f_s(\mathbf{x}) \odot f_e(\mathbf{x}) \\ &= f_s(\mathbf{x}) \cdot f_e(\mathbf{x}) - \mathcal{H}\big(f_s(\mathbf{x})\big) \cdot \mathcal{H}\big(f_e(\mathbf{x})\big) \end{aligned}$$

where $\mathcal{H}\big(f(\mathbf{x})\big)$ is the Hilbert transform of signal $f(\mathbf{x})$, and the smooth and local energy components are derived below, respectively,

$$f_s(\mathbf{x}) = \frac{\check{f}(\mathbf{x}) \cdot f_e(\mathbf{x}) + \mathcal{H}\big(\check{f}(\mathbf{x})\big) \cdot \mathcal{H}\big(f_e(\mathbf{x})\big)}{f_e^2(\mathbf{x}) + \mathcal{H}^2\big(f_e(\mathbf{x})\big)} \qquad f_e(\mathbf{x}) = \sqrt{\check{f}^2(\mathbf{x}) + \mathcal{H}^2\big(\check{f}(\mathbf{x})\big)}$$

The reconstruction of $f(\mathbf{x})$ with the Direct Current (DC) component nullified is $\check{f}(\mathbf{x})$.

Time-frequency analysis

The power spectrum of $f_e(\mathbf{x})$ is calculated for every pixel location. The metric $g(\mathbf{x})$ summates over the normalized powers which peak along the period axis. Considering only the maximal values of periodic power, there is no need for any windowing function, which in practice may dampen the values we seek and increase the minimum length of the time window n required.

Let $\mathcal{P} = \{p_1(\mathbf{x}), p_2(\mathbf{x}), \cdots, p_\tau(\mathbf{x})\}$ be the sorted set of peaks of normalized power in descending order where $\tau < \lfloor n/2 \rfloor$. Then the motion classification metric is defined as follows:

$$g(\mathbf{x}) = \sum_{i=1}^{k} p_i(\mathbf{x}), \quad k \le \tau$$

$g(\mathbf{x})$ 0 α β 1

irregular motion — periodic motion

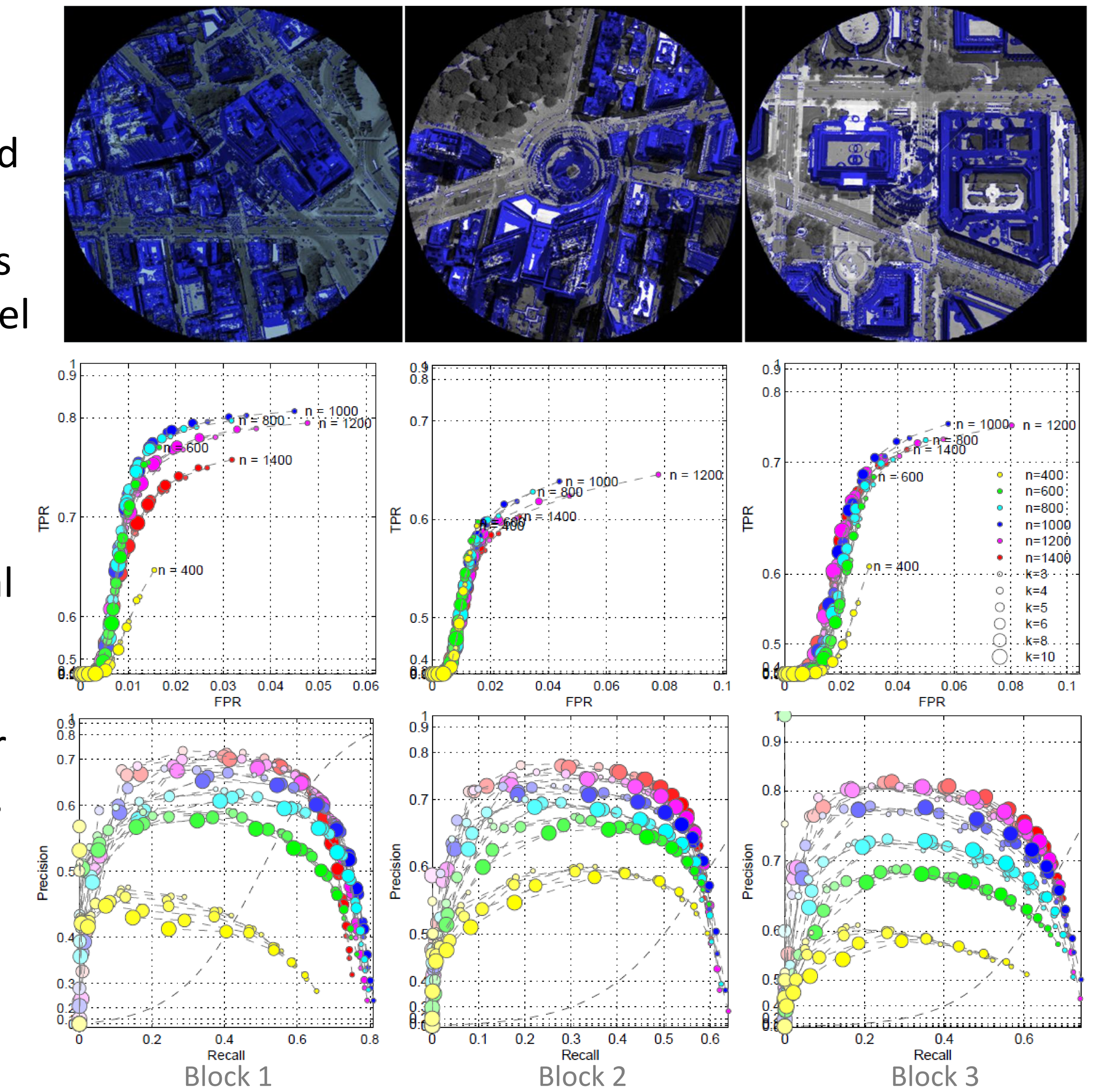
The typical ranges of the metric's parameters are $3 \le k \le 10$, $\alpha \in [0.05, 0.30]$, and $\beta \in [0.50, 0.90]$. The time window n should observe at least a single cycle of precession.

Experiments and results

Results are shown for a qualitative view structural motion detection using artificial data (first row) and for independent motion detection using proprietary, real WAMI showing ROC and PR curves (color graphs). The run-time of the method per pixel is $O(n \log n)$.

In the first row, regions undergoing precession are tinted blue with $\beta = 0.5$. Like many pixel-based detectors, this does not guarantee detection for region interiors due to surface homogeneity. In real data, surfaces glean in the sensor appearing inhomogeneous over time.

For WAMI, precision increases with n and peaks for specific combinations of k and α, across all blocks. Precision confidence for all curves increases with the amount of independent motion. Gains for all sequences diminish once two cycles of precession have been observed ($n = 600$).



This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344.

LLNL-POST-557781 UNCLASSIFIED